



2-Day Course – Spatial Modeling with Geostatistics

**Prof. Michael J. Pyrcz, Ph.D., P.Eng.
Associate Professor**

**Hildebrand Department of Petroleum & Geosystems Engineering
University of Texas at Austin**

**Bureau of Economic Geology, Jackson School of Geosciences
University of Texas at Austin**

**“In two days, what a geoscientists needs to know about geostatistics, and
workflows to get you started with applying geostatistics to impact your work.”**

Spatial Modeling with Geostatistics



Spatial Estimation

Lecture outline . . .

- Trend + Residual Method
- Trend Modeling
- Spatial Estimation

Prerequisites

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

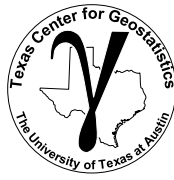
Spatial Estimation

Stochastic Simulation

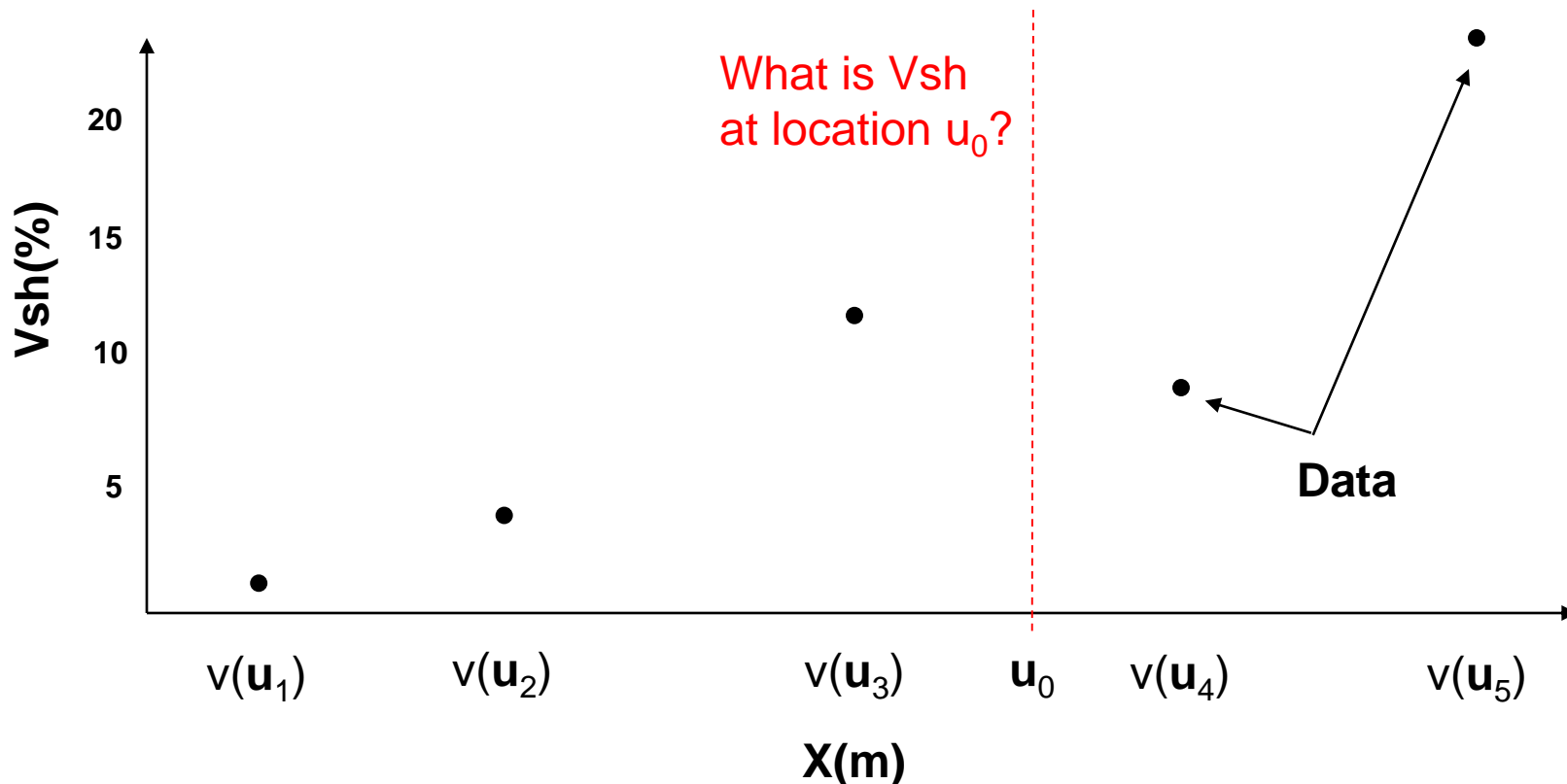
Uncertainty Management

Machine Learning

Estimation with Nonstationarity



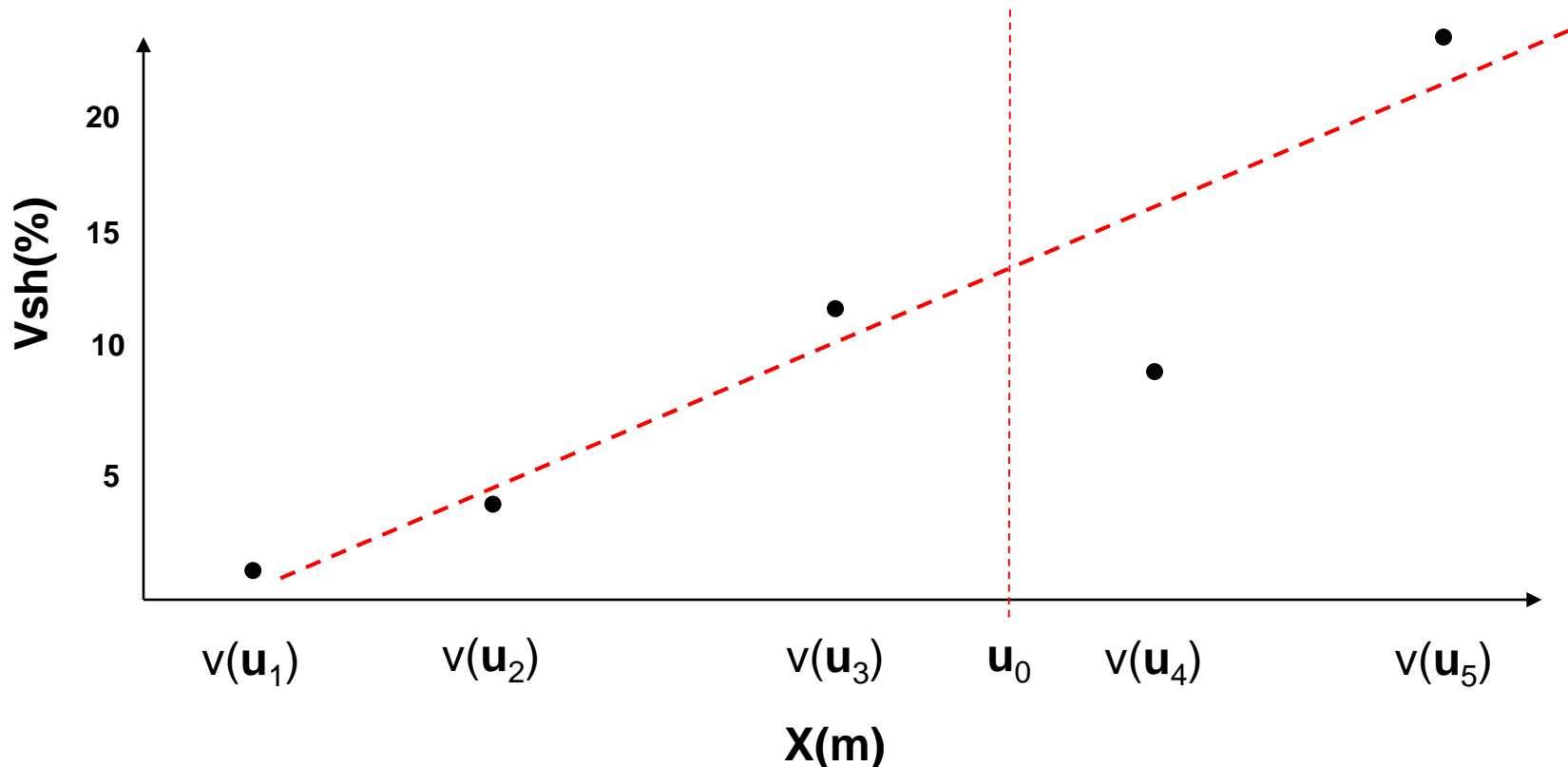
- Geostatistical spatial estimation methods will make an assumption concerning stationarity
 - In the presence of significant nonstationarity we would not rely 100% for spatial estimation on data + spatial continuity model



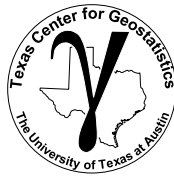
Estimation with Nonstationarity



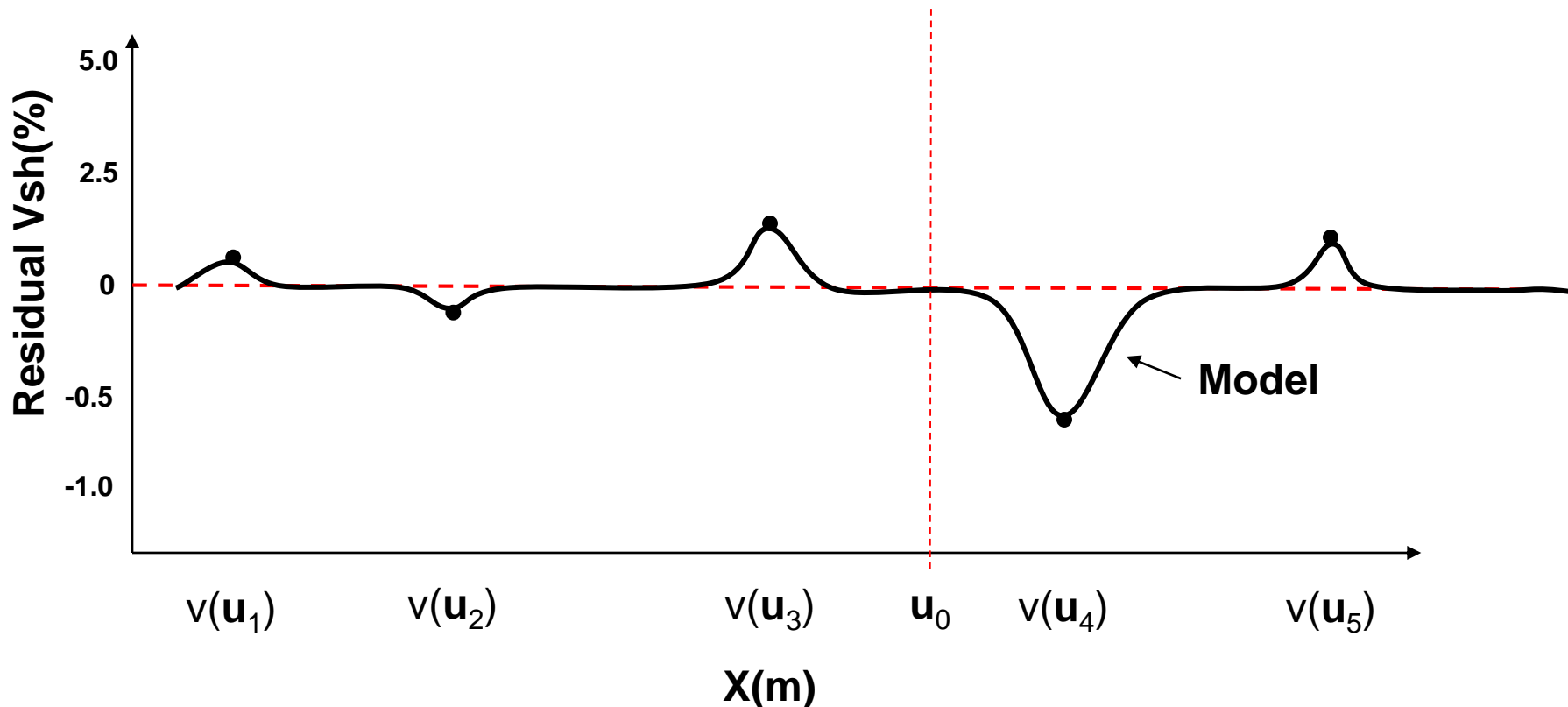
- Geostatistical spatial estimation methods will make an assumption concerning stationarity
 - If we observe a trend, we should model the trend.



Estimation with Nonstationarity



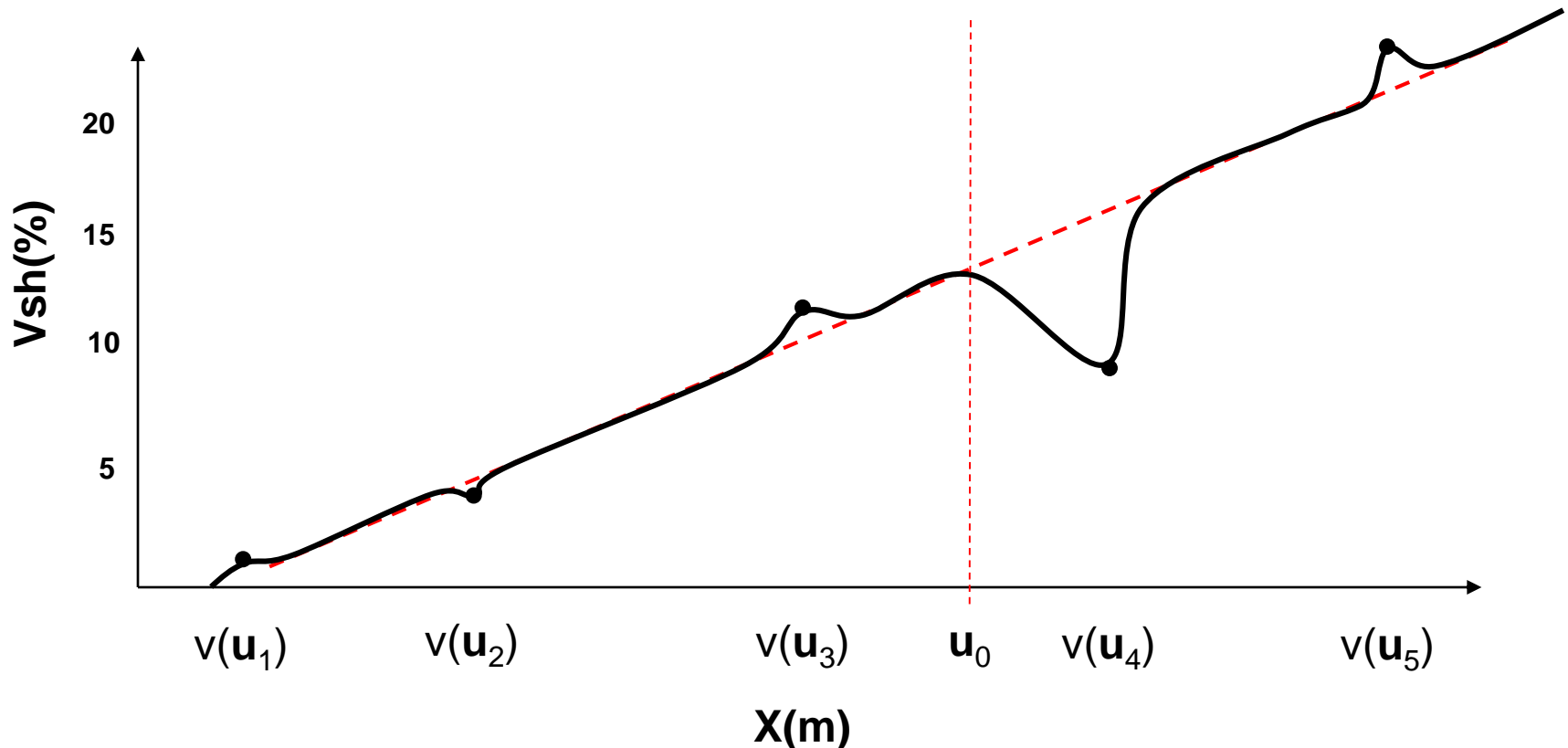
- Geostatistical spatial estimation methods will make an assumption concerning stationarity
 - Then model the residuals.



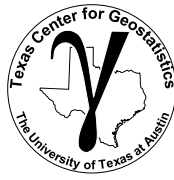
Estimation with Nonstationarity



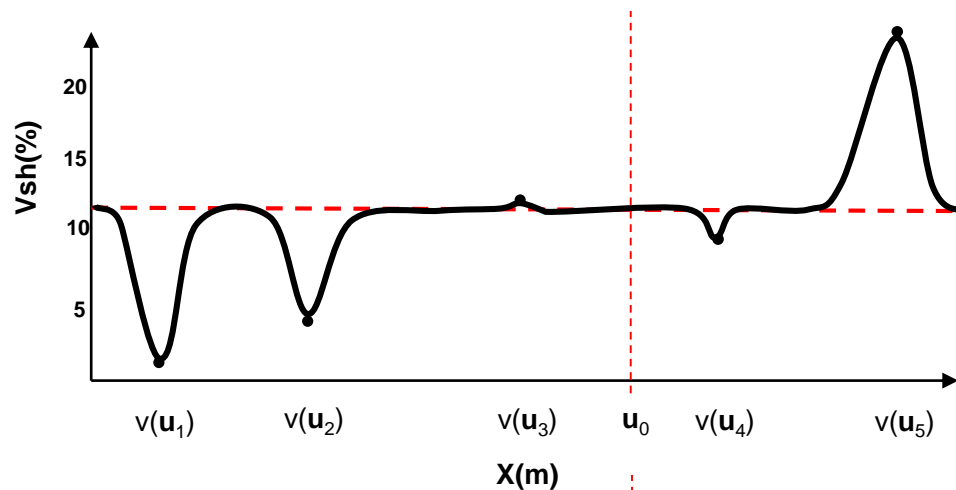
- Geostatistical spatial estimation methods will make an assumption concerning stationarity
 - After modeling, add the trend back to the modelled residuals



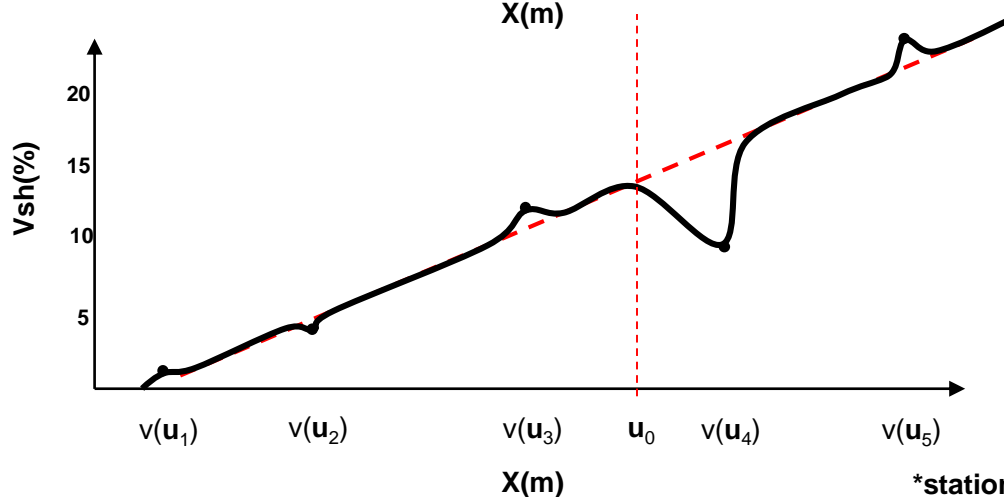
Estimation with Nonstationarity



- How bad could it be if we did not model a trend?
- Geostatistical estimation would assume stationarity* and away from data we would estimate with the global mean (simple kriging)!



**Model with
stationary
mean + data.**



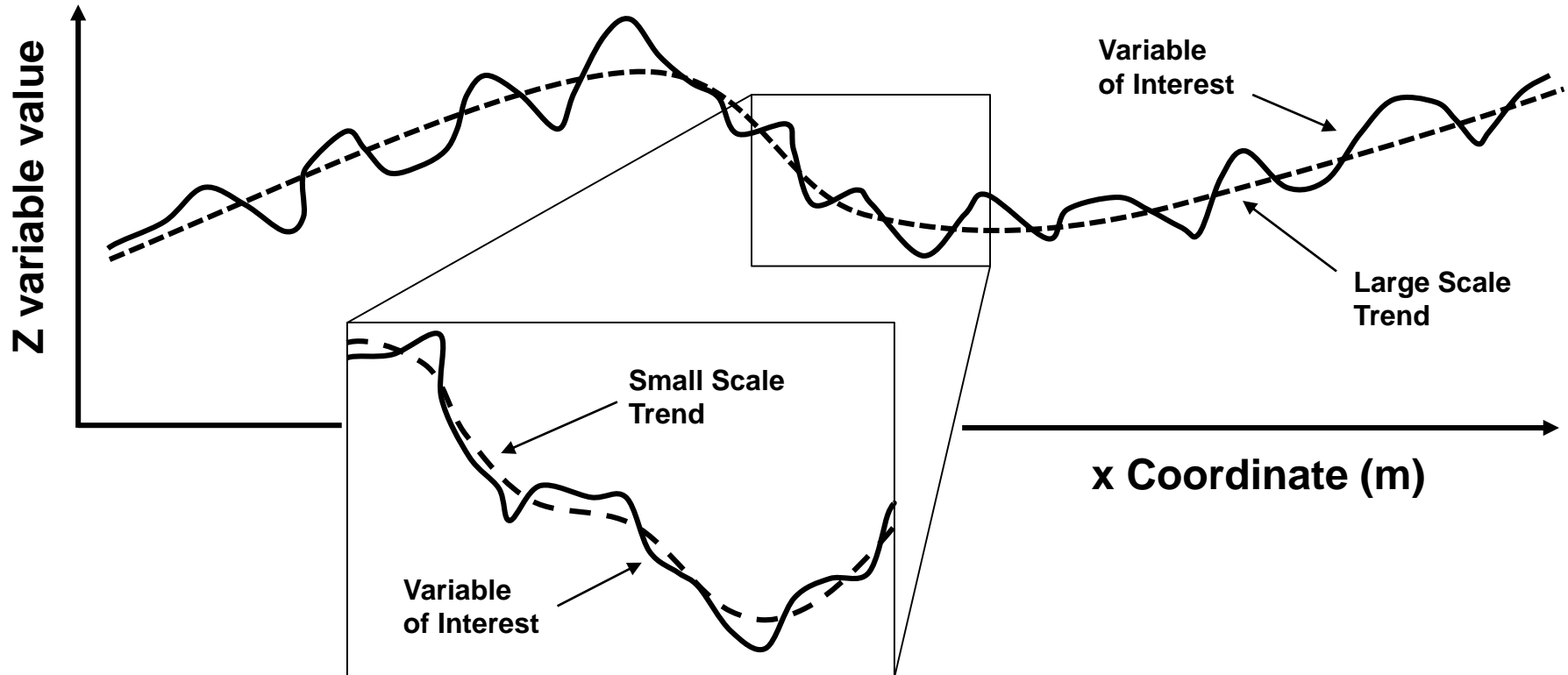
**Model with
mean trend model
and residual + data.**

*stationarity decision depends on type of method.

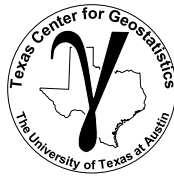
Estimation with Nonstationarity



- Trend Modeling
 - We must identify and model trends / nonstationarities



Estimation with Nonstationarity



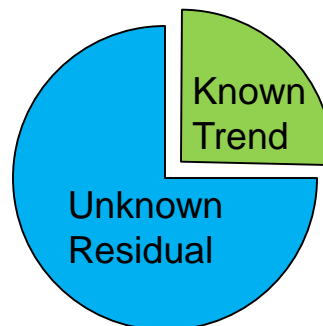
- Any variance in the trend is removed from the residual:

$$\sigma_X^2 = \sigma_{X_t}^2 + \sigma_{X_r}^2 + 2C_{X_t, X_r}$$

- if the $X \perp Y$, $C_{X_t, X_r} = 0$

$$\sigma_{X_r}^2 = \sigma_X^2 - \sigma_{X_t}^2$$

- So if σ_X^2 is the total variance (variability), and $\sigma_{X_t}^2$ is the variability that is deterministically modelled, treated as known, and $\sigma_{X_r}^2$ is the component of the variability that is treated as unknown.
- Result: the more variability explained by the trend the less variability that remains as uncertain.

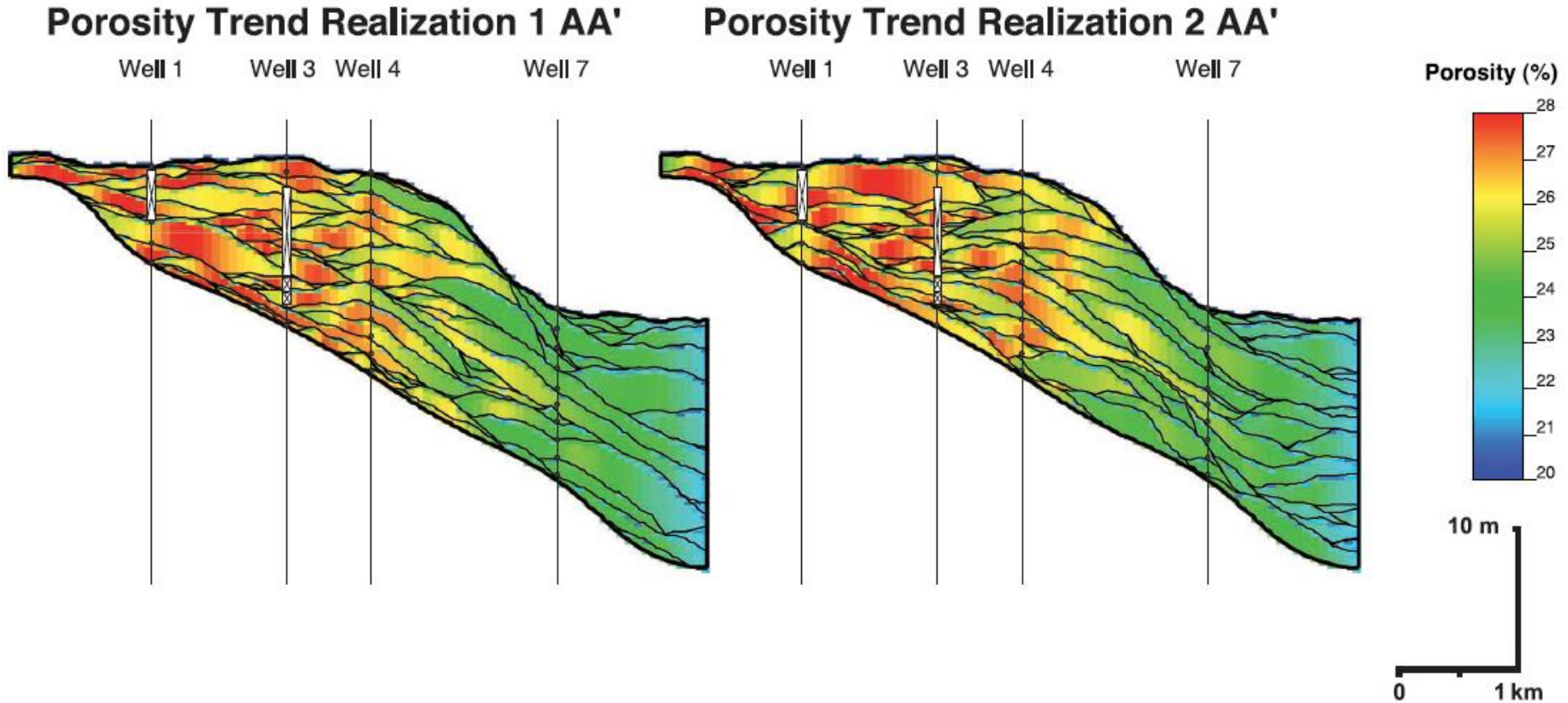


Definition Deterministic Model



- Model that assumes perfect knowledge, without uncertainty
- Based on knowledge of the phenomenon or trend fitting to data
- Most subsurface models have a deterministic component (trend) to capture expert knowledge and to provide a stationary residual for geostatistical modeling.

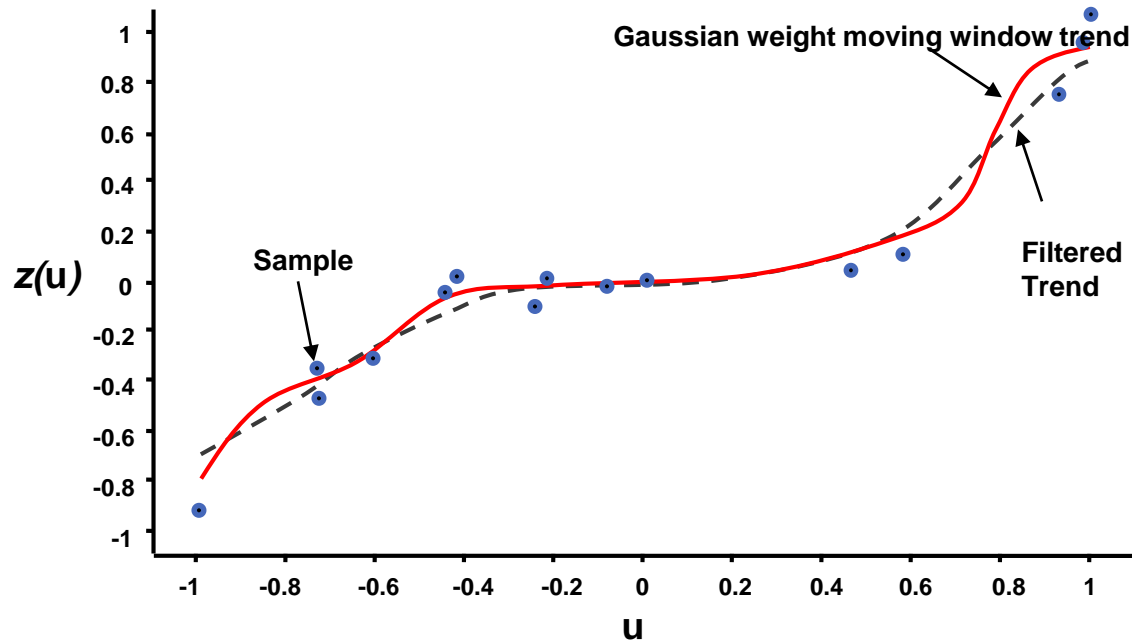
Trend Modeling



- Trend models:
 - Tend to be smooth, based on data and interpretation
 - May be complicated (see above)
 - Parameterized by vertical proportion curves (see below) and areal trend maps

Trend Modeling Workflow

- How to calculate a trend model:
 - Moving window average of the available data
 - Weighting scheme within the window
 - » Uniform weights can cause discontinuities
 - » Reduce weight at edges of moving window to reduce discontinuities (e.g. Gaussian weights).



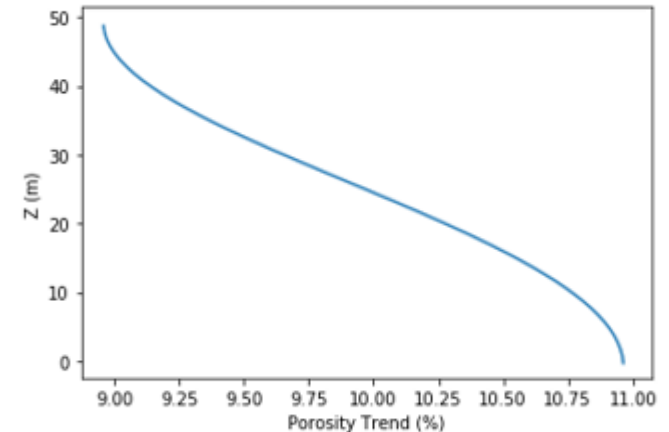
Trend Modeling Workflow

- How to calculate a trend model:
 - Calculate the 2D areal trend by interpolating over vertically averaged wells.
 - Calculate the 1D vertical trend by averaging wells layer-by-layer
 - Combine the 1D vertical and the 2D areal trends, one method is:

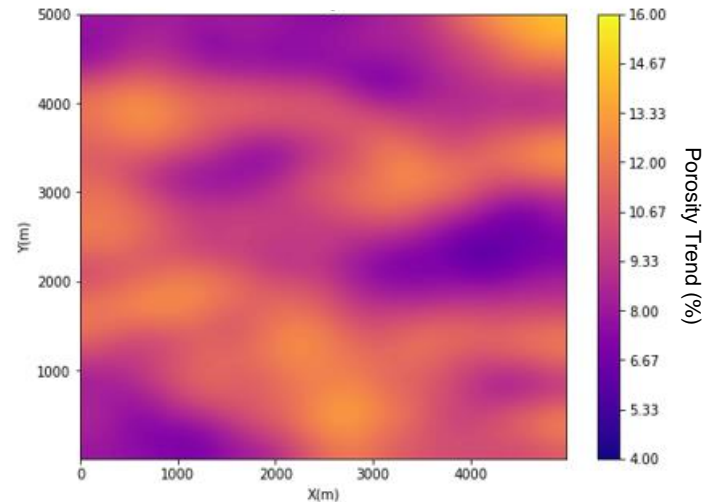
$$\bar{X}(x, y, z) = \bar{X}(z) \cdot \frac{\bar{X}(x, y)}{\bar{X}}$$

3D Trend \rightarrow $\bar{X}(x, y, z)$ 1D Vertical Trend \rightarrow $\bar{X}(z)$ 2D Areal Trend \rightarrow $\bar{X}(x, y)$ Global Mean \rightarrow \bar{X}

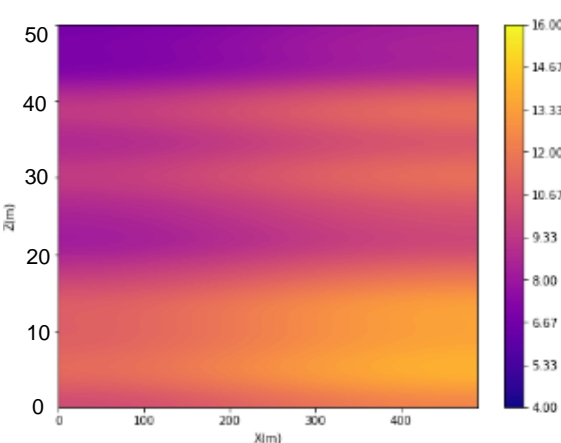
1D Vertical Porosity Trend



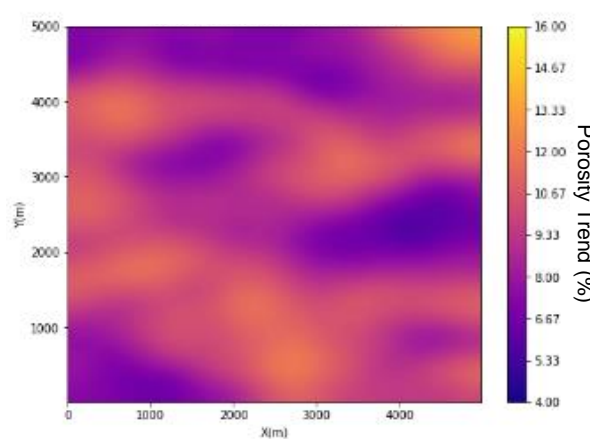
2D Areal Porosity Trend



3D Porosity Trend, y = 2500m



3D Porosity Trend, z = 10m

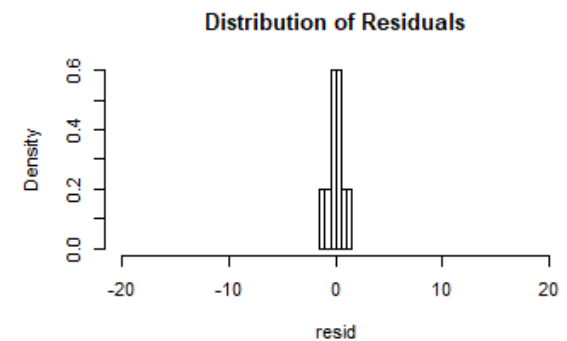
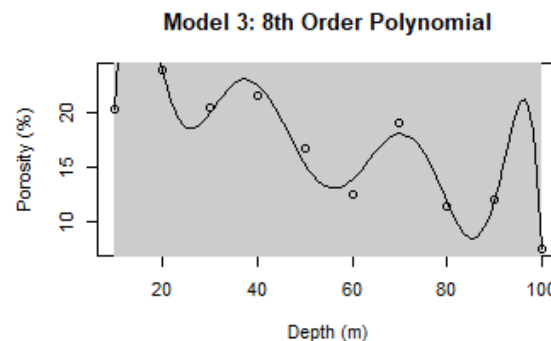
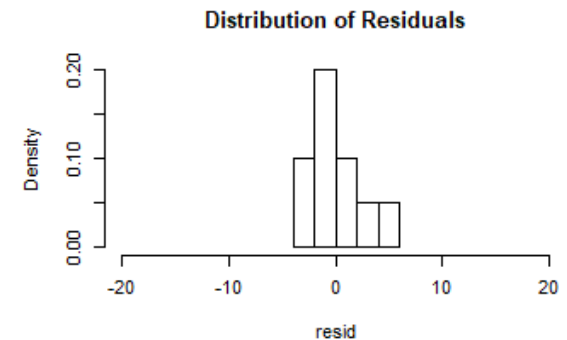
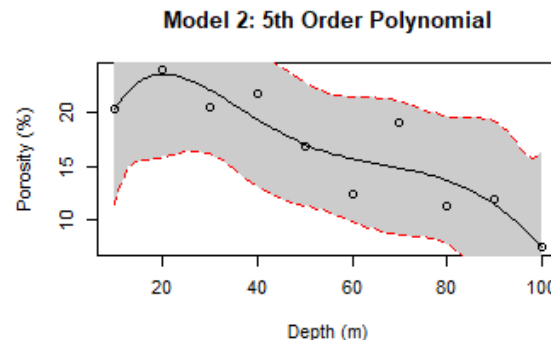
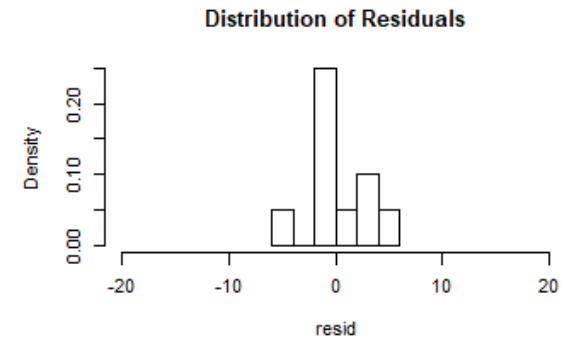
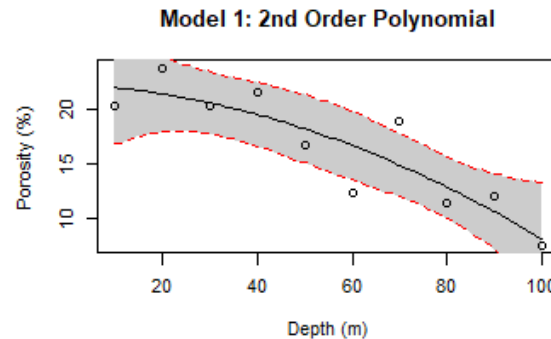


Trend Definition

- Observation of nonstationarity in any statistic, metric of interest
 - e.g. Proportions of facies, mean of porosity
- A model of the nonstationarity in any statistic, metric of interest
 - e.g. 3D model of locally variable facies proportion and mean porosity
- Typically modeled with support of data and expert knowledge in a deterministic manner (without uncertainty).
 - May use multiple trend scenarios to account for trend uncertainty

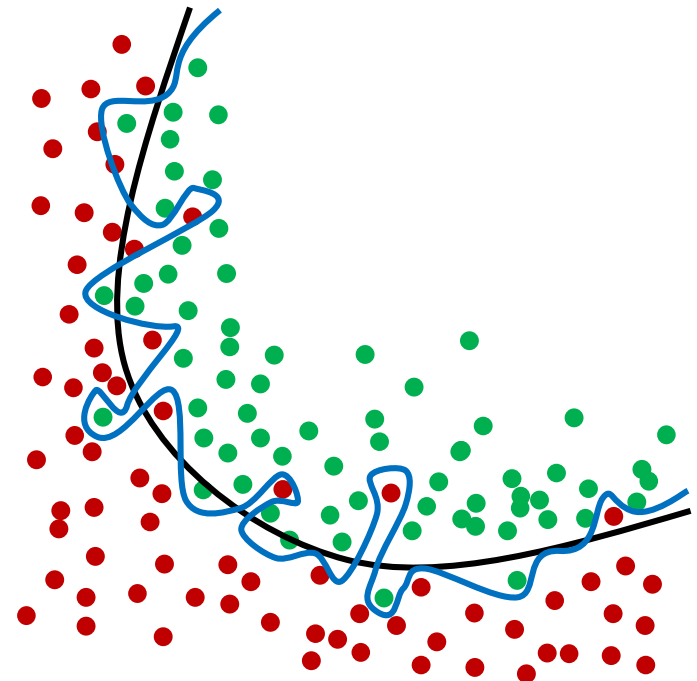
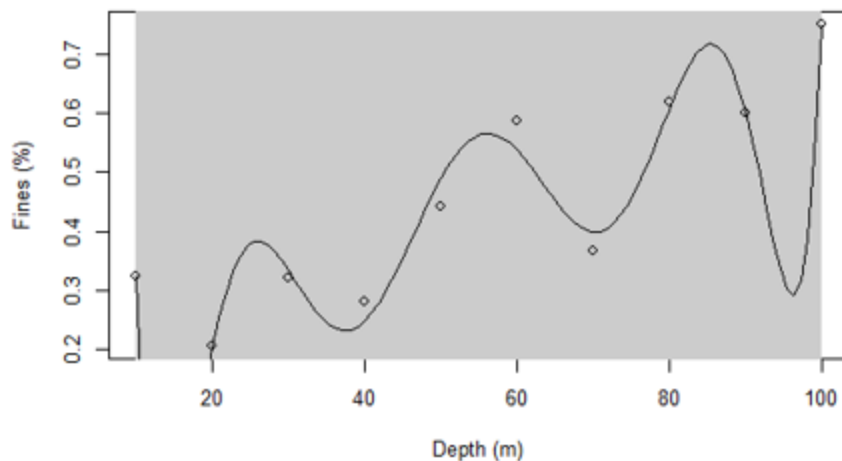
Overfitting Trend Models

- 1D polygonal fit porosity (%) trend with respect to depth
- As model complexity increases the model is more flexible to fit the data
- The residual (unexplained variance goes down)
- The model confidence interval increases
- The predictions are more extreme in extrapolation



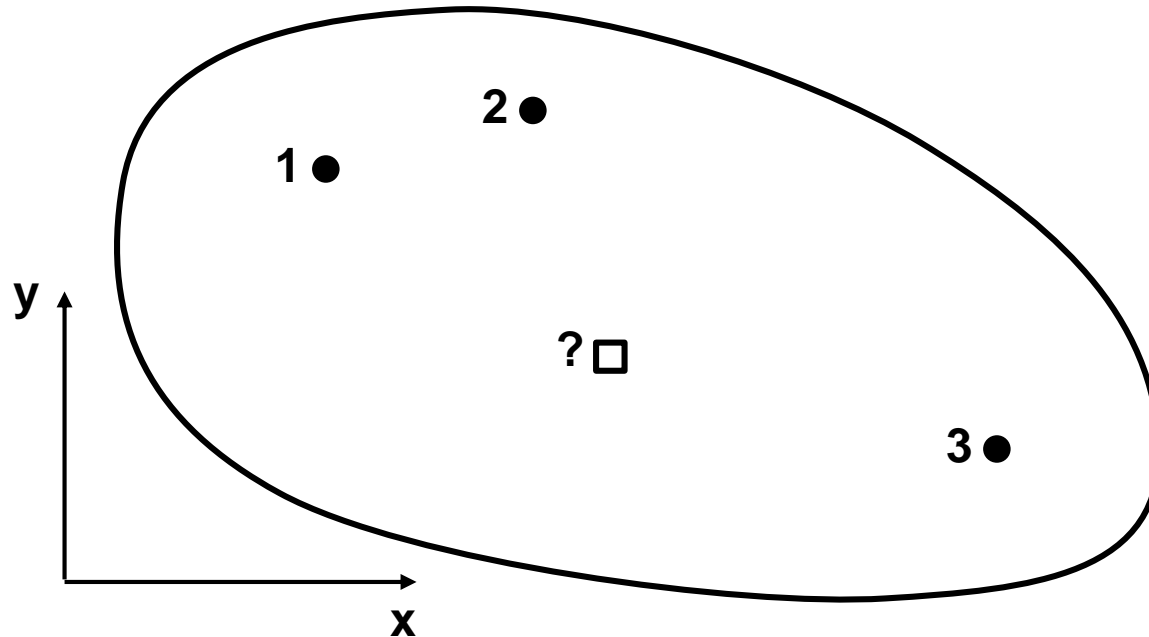
Definition of Overfitting

- Overly complicated model to explain “idiosyncrasies” of the data, capturing data noise in the model
- More parameters than can be justified with the data
- Results in likely very high error away from the data
- But, results in low residual variance!
- High R^2
- Very accurate with the training data!
 - Claim you know more than you actually do!



Spatial Estimation

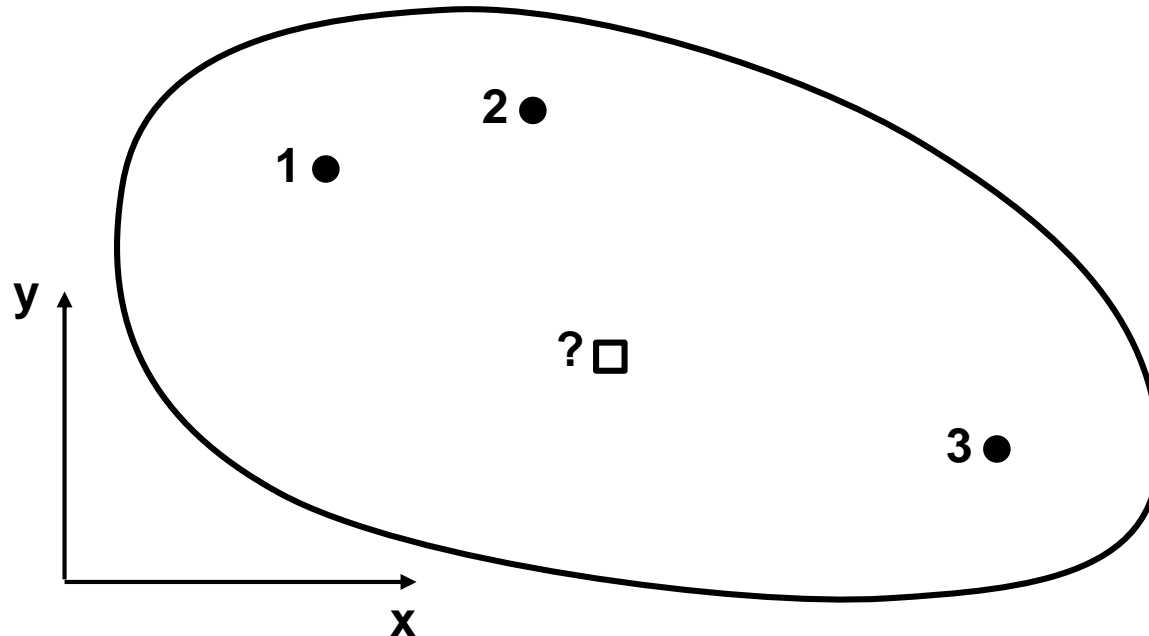
- Consider the case of estimating at some unsampled location:



- How would you do this given data, $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$, and $z(\mathbf{u}_3)$?

Spatial Estimation

- Consider the case of estimating at some unsampled location:



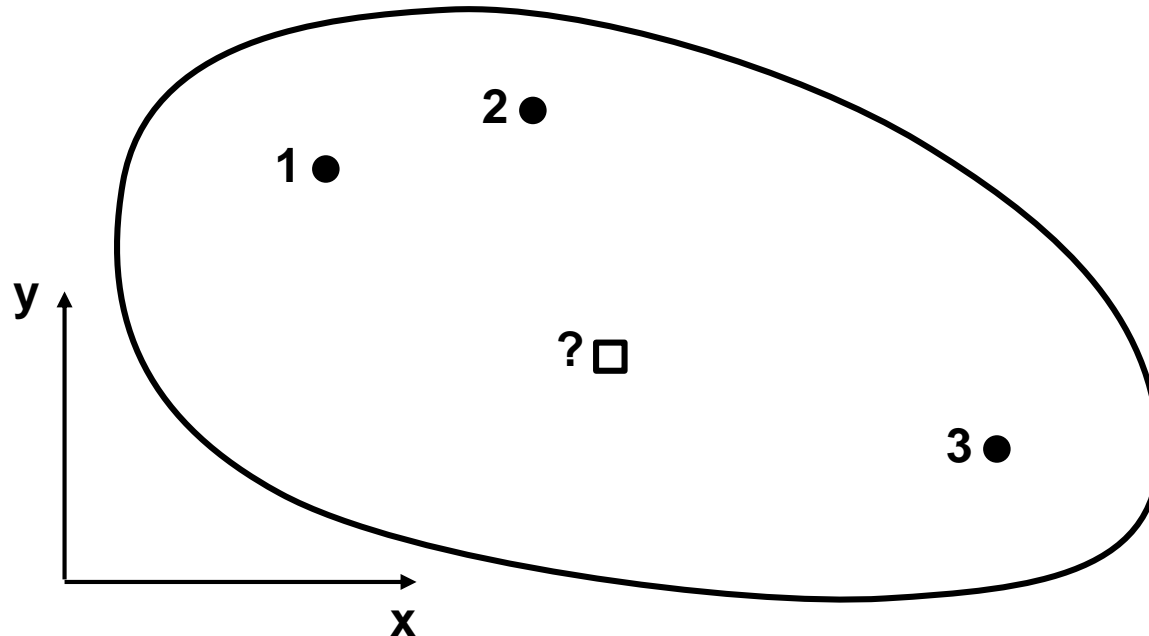
- How would you do this given data, $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$, and $z(\mathbf{u}_3)$?

$$z^*(\mathbf{u}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{u}_{\alpha}) + \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha} \right) m_z$$

**Unbiasedness
Constraint
Weights sum to 1.0.**

Spatial Estimation

- Consider the case of estimating at some unsampled location:



- How would you do this given data, $z(\mathbf{u}_1)$, $z(\mathbf{u}_2)$, and $z(\mathbf{u}_3)$?

$$z^*(\mathbf{u}_0) - m_z(\mathbf{u}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} (z(\mathbf{u}_{\alpha}) - m_z(\mathbf{u}_{\alpha}))$$

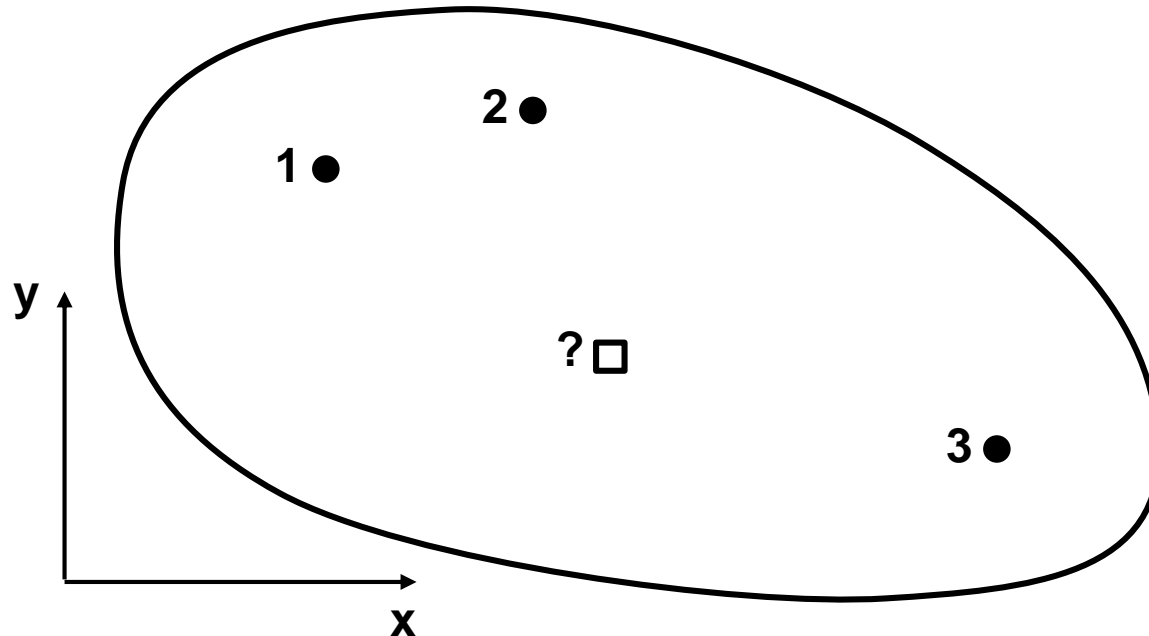
In the case where the mean is non-stationary.

$$Y = Z - m, \quad y^*(\mathbf{u}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} y(\mathbf{u}_{\alpha})$$

Simplified with residual, y .

Spatial Estimation

- Consider the case of estimating at some unsampled location:



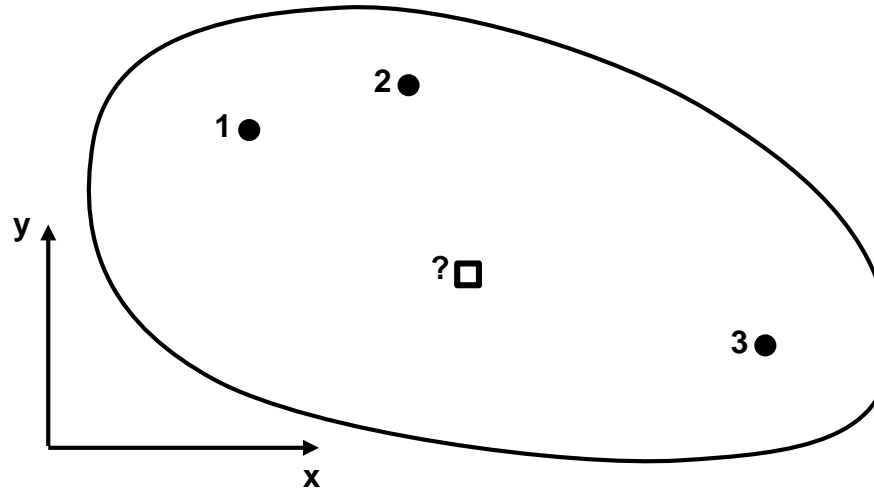
- Linear weighted, sound good. How do we get the weights? $\lambda_\alpha, \alpha = 1, \dots, n$

$$y^*(u_0) = \sum_{\alpha=1}^n \lambda_\alpha y(u_\alpha)$$

**Simplified with residual, y .
we have subtracted a trend.**

Spatial Estimation

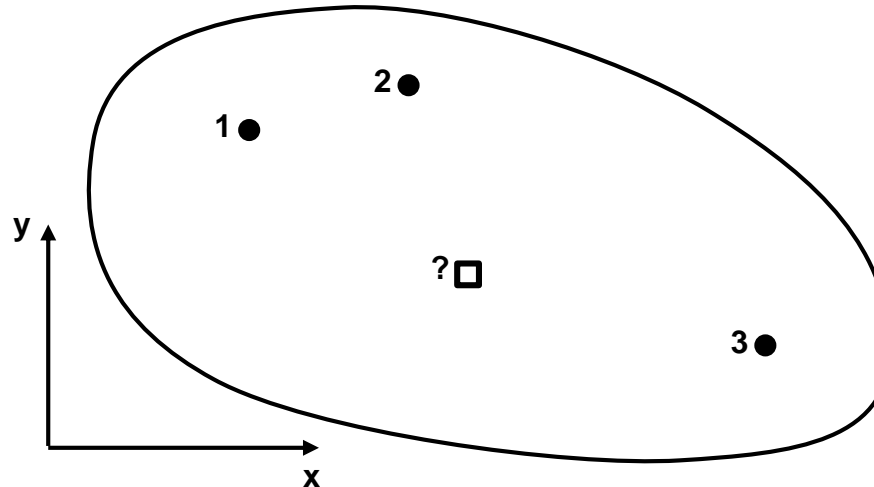
- Consider the case of estimating at some unsampled location:



- Linear weighted, sound good. How do we get the weights? $\lambda_\alpha, \alpha = 1, \dots, n$
- Equal weighted / average? $\lambda_\alpha = 1/n$ **Equal weight
average of data**
- What's wrong with that?

Spatial Estimation

- Consider the case of estimating at some unsampled location:



- How do we get the weights? $\lambda_\alpha, \alpha = 1, \dots, n$

- Inverse distance?

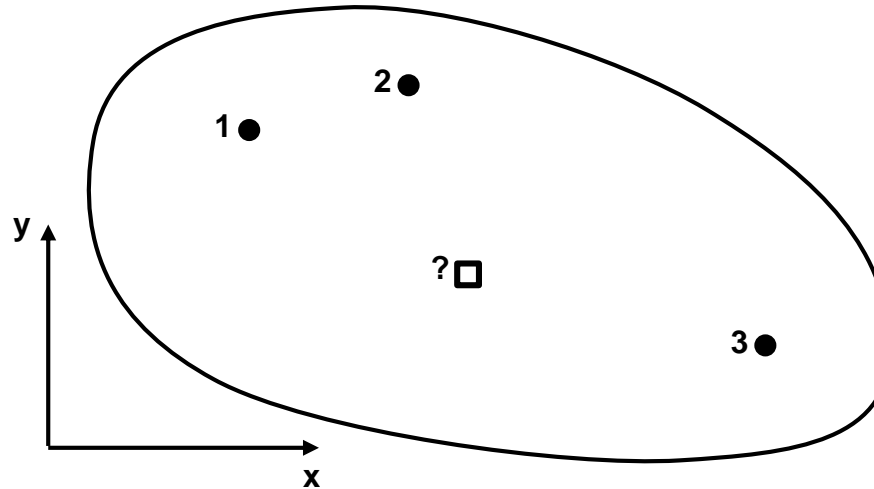
$$\lambda_\alpha = \frac{1}{\text{dist}(\mathbf{u}_0, \mathbf{u}_\alpha)^p} / \sum_{\alpha=1}^n \lambda_\alpha$$

**Inverse distance to power
standardized so weights
sum to 1.0.**

- What's wrong with that?

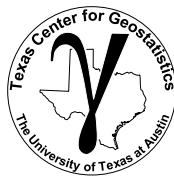
Spatial Estimation

- Consider the case of estimating at some unsampled location:



- How do we get the weights? $\lambda_{\alpha}, \alpha = 1, \dots, n$
- It would be great to use weight that account for closeness (spatial correlation > distance alone), redundancy (once again with spatial correlation).
- How can we do that?

Derivation of Simple Kriging Equations



- Consider a linear estimator:

$$Y^*(\mathbf{u}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Y(\mathbf{u}_{\alpha})$$

where $Y(\mathbf{u}_i)$ are the residual data (data values minus the mean) and $Y^*(\mathbf{u}_i)$ is the estimate (add the mean back in when we are finished)

- The **error variance** is defined as:

Stationary Mean, Variogram

$$E\{Y\} = 0$$

$$2\gamma(\mathbf{h}) = E\left\{[Y(\mathbf{u}) - Y(\mathbf{u} + \mathbf{h})]^2\right\}$$

$$\begin{aligned} E\{[Y^*(\mathbf{u}) - Y(\mathbf{u})]^2\} &= E\{[Y^*(\mathbf{u})]^2\} - 2 \cdot E\{Y^*(\mathbf{u}) \cdot Y(\mathbf{u})\} + E\{[Y(\mathbf{u})]^2\} \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E\{Y(\mathbf{u}_i) \cdot Y(\mathbf{u}_j)\} - 2 \cdot \sum_{i=1}^n \lambda_i E\{Y(\mathbf{u}) \cdot Y(\mathbf{u}_i)\} + C(0) \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{u}_i, \mathbf{u}_j) - 2 \cdot \sum_{i=1}^n \lambda_i C(\mathbf{u}, \mathbf{u}_i) + C(0) \end{aligned}$$

redundancy
closeness
variance

More Derivation

- Optimal weights $\lambda_i, i = 1, \dots, n$ may be determined by taking partial derivatives of the error variance w.r.t. the weights

$$\frac{\partial[\quad]}{\partial \lambda_i} = \sum_{j=1}^n \lambda_j C(\mathbf{u}_i, \mathbf{u}_i) - 2 \cdot C(\mathbf{u}, \mathbf{u}_i), i = 1, \dots, n$$

and setting them to zero

$$\sum_{j=1}^n \lambda_j C(\mathbf{u}_i, \mathbf{u}_i) = \cdot C(\mathbf{u}, \mathbf{u}_i), i = 1, \dots, n$$

- This system of n equations with n unknown weights is the simple kriging (SK) system



Kriging Definition

- Estimation approach that relies on linear weights that account for spatial continuity, data closeness and redundancy.
- Weights are unbiased and minimize the estimation variance.

Simple Kriging: Some Details

There are three equations to determine the three weights:

$$\lambda_1 \cdot C(\mathbf{u}_1, \mathbf{u}_1) + \lambda_2 \cdot C(\mathbf{u}_1, \mathbf{u}_2) + \lambda_3 \cdot C(\mathbf{u}_1, \mathbf{u}_3) = C(\mathbf{u}_0, \mathbf{u}_1)$$

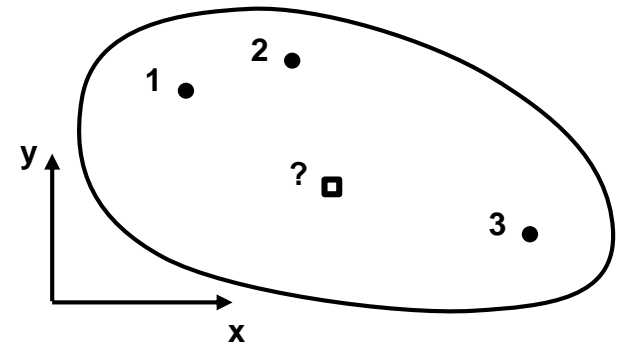
$$\lambda_1 \cdot C(\mathbf{u}_2, \mathbf{u}_1) + \lambda_2 \cdot C(\mathbf{u}_2, \mathbf{u}_2) + \lambda_3 \cdot C(\mathbf{u}_2, \mathbf{u}_3) = C(\mathbf{u}_0, \mathbf{u}_2)$$

$$\lambda_1 \cdot C(\mathbf{u}_3, \mathbf{u}_1) + \lambda_2 \cdot C(\mathbf{u}_3, \mathbf{u}_2) + \lambda_3 \cdot C(\mathbf{u}_3, \mathbf{u}_3) = C(\mathbf{u}_0, \mathbf{u}_3)$$

In matrix notation: Recall that $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$

$$\begin{bmatrix} C(\mathbf{u}_1, \mathbf{u}_1) & C(\mathbf{u}_1, \mathbf{u}_2) & C(\mathbf{u}_1, \mathbf{u}_3) \\ C(\mathbf{u}_2, \mathbf{u}_1) & C(\mathbf{u}_2, \mathbf{u}_2) & C(\mathbf{u}_2, \mathbf{u}_3) \\ C(\mathbf{u}_3, \mathbf{u}_1) & C(\mathbf{u}_3, \mathbf{u}_2) & C(\mathbf{u}_3, \mathbf{u}_3) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C(\mathbf{u}_0, \mathbf{u}_1) \\ C(\mathbf{u}_0, \mathbf{u}_2) \\ C(\mathbf{u}_0, \mathbf{u}_3) \end{bmatrix}$$

redundancy **closeness**



Properties of Simple Kriging



- Solution exists and is unique if matrix $[C(v_i, v_j)]$ is positive definite
- Kriging estimator is unbiased: $E\{[Z_V - Z_K^*]\} = 0$
- Minimum error variance estimator (just try to pick weights, you won't bet it)
- Best Linear Unbiased Estimator
- Provides a measure of the estimation (or kriging) variance (uncertainty in the estimate):

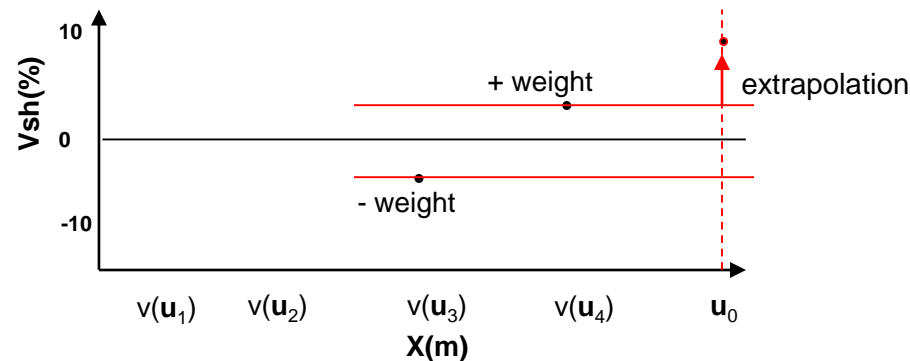
$$\sigma_E^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u} - \mathbf{u}_{\alpha}) \quad \sigma_E^2 \rightarrow [0, \sigma_x^2]$$

More Properties

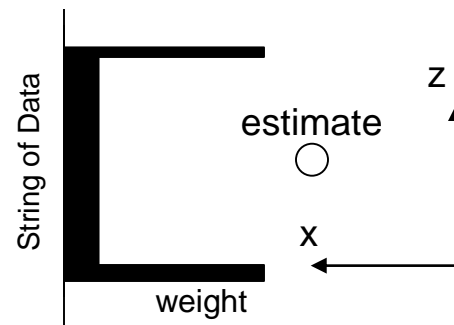
- Exact interpolator: at data location
- Kriging variance can be calculated before getting the sample information, homoscedastic!
- Kriging takes into account:
 - distance of the information: $C(\mathbf{u}, \mathbf{u}_i)$
 - configuration of the data: $C(\mathbf{u}_i, \mathbf{u}_j)$
 - structural continuity of the variable being considered: $C(\mathbf{h})$
- The smoothing effect of kriging can be forecast – we will return to this with simulation.
- Kriging theory is part of the probabilistic theory of *projectors*: orthogonal projection onto space of linear combinations of the n data (Hilbert space)

More Properties

- Outside range of ant data, simple kriging weights are all equal 0.0. The best estimate is the provided mean!
- Screened data will sometimes have negative weights! This allows kriging to extrapolate.



- Strings of data will have an artifact known as the string effect.



Simple Kriging Hands-on

Simple Kriging Demonstration

Michael Pyroz, Geostatistics at Petroleum and Geosystems Engineering, University of Texas at Austin (mpyroz@austin.utexas.edu)

File at Excel/Simple_Kriging_Demo.xls

1. Data and Estimate Locations and Values

Point	x	y	value	residual
1	60	80	0.1	-0.040
2	25	50	0.12	-0.020
3	80	10	0.2	0.060
unknown	50	50		
mean			0.140	

2. Distance Matrix

0.00	46.10	72.80	31.62
46.10	0.00	68.01	25.00
72.80	68.01	0.00	50.00

3. Variogram Model

Nugget	0
Spherical	1
Range	300

4. Variogram Matrix

0.000	0.229	0.357	0.158
0.229	0.000	0.334	0.125
0.357	0.334	0.000	0.248

5. Covariance Matrix

1.000	0.771	0.643	0.842
0.771	1.000	0.666	0.875
0.643	0.666	1.000	0.752

6. Inverse Left Side

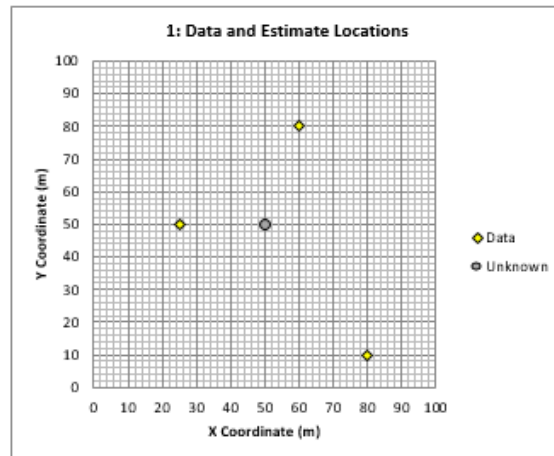
2.667	-1.644	-0.621
-1.644	2.810	-0.813
-0.621	-0.813	1.941

7. Weights

0.341
0.462
0.225
Sum Weights
Mean Weight

8. Kriging Results

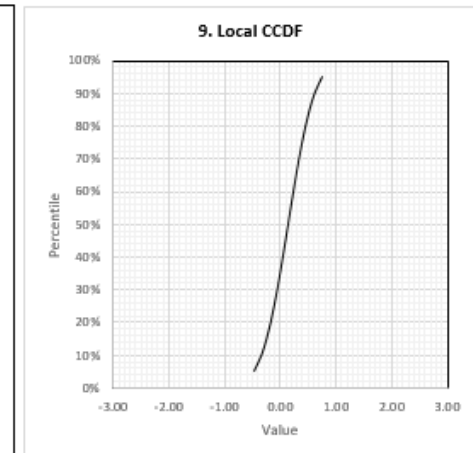
Kriging Estimate	0.131
Kriging Variance	0.139



Legend

Information
User Input
Calculation
Redundancy Measure
Closeness Measures

p-value	F ⁻¹ (p)
5%	-0.48
10%	-0.35
15%	-0.26
20%	-0.18
25%	-0.12
30%	-0.06
35%	-0.01
40%	0.04
45%	0.08
50%	0.13
55%	0.18
60%	0.22
65%	0.27
70%	0.33
75%	0.38
80%	0.44
85%	0.52
90%	0.61
95%	0.74



Description

This sheet provides an illustration of Simple Kriging at a single estimated location.

- Step 1: Input the data locations and values, the unknown simulated location. At any point these locations and values may be changed to observed their influence on the simulation.
- Step 2: The distance matrix is automatically calculated, that is the distance between the data and the unknown locations.
- Step 3: Enter the model of spatial continuity in the form of an isotropic spherical variogram and nugget effect (contributions should sum to one). This model may be changed at any time to observed sensitivities to spatial continuity.
- Step 4: Variogram matrix is calculated by applying the distance matrix to the isotropic variogram model.
- Step 5: Covariance matrix is calculated by subtracting the variogram from the variance (1 for standard normal distribution). This is applied to improve numerical stability as a diagonally dominant matrix is more readily invertible.
- Step 6: The left hand side of the covariance matrix is inverted.
- Step 7: The inverted left handside matrix is multiplied by the right hand side matrix to calculate the simple kriging weights.
- Step 8: The kriging estimate and kriging variance are calculated with the weights and covariances.
- Step 9: With the Gaussian assumption the complete local conditional cumulative distribution function is available.

Simple Kriging Hands-on



- Some ideas for experimenting with simple kriging. Do the following and pay attention to the weights, the estimate and the estimation variance.
 1. Set points 1 and 2 closer together.
 2. Put point 1 behind point 2 to create screening.
 3. Put all points outside the range.
 4. See the range very large.

Ordinary Kriging: Some Details



Add the constraint of : $\sum_{\alpha=1}^n \lambda_{\alpha} = 1.0$

$$\begin{bmatrix} C(\mathbf{u}_1, \mathbf{u}_1) & C(\mathbf{u}_1, \mathbf{u}_2) & C(\mathbf{u}_1, \mathbf{u}_3) & 1 \\ C(\mathbf{u}_2, \mathbf{u}_1) & C(\mathbf{u}_2, \mathbf{u}_2) & C(\mathbf{u}_2, \mathbf{u}_3) & 1 \\ C(\mathbf{u}_3, \mathbf{u}_1) & C(\mathbf{u}_3, \mathbf{u}_2) & C(\mathbf{u}_3, \mathbf{u}_3) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu \end{bmatrix} = \begin{bmatrix} C(\mathbf{u}_0, \mathbf{u}_1) \\ C(\mathbf{u}_0, \mathbf{u}_2) \\ C(\mathbf{u}_0, \mathbf{u}_3) \\ 1 \end{bmatrix}$$

In matrix notation: Recall that $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$

$$z^*(\mathbf{u}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{u}_{\alpha}) + \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha} \right) m_z$$

Note: In the original image, a red diagonal line is drawn through the term $(1 - \sum \lambda_{\alpha})$ and the mean m_z , with a red '0' above the line, indicating that this term is zero in ordinary kriging.

With ordinary kriging the mean does not need to be known. Ordinary kriging estimates the mean locally!

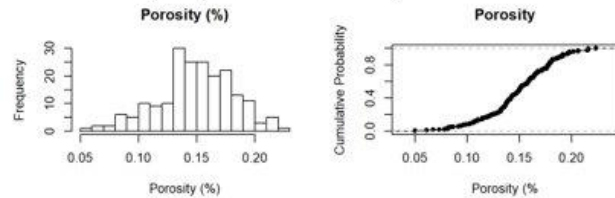
Kriging

- Kriging is a procedure for constructing a minimum error variance linear estimate at a location where the true value is unknown
- The main controls on the kriging weights are:
 - closeness of the data to the location being estimated
 - redundancy between the data
 - the variogram
- Simple Kriging (SK) does not constrain the weights and works with the residual from the mean
- Ordinary Kriging (OK) constrains the sum of the weights to be 1.0, therefore, the mean does not need to be known
- There are many different types of kriging
- Two implicit assumptions are stationarity (work around with different types of kriging) and ergodicity (more slippery)
- Kriging is not used directly for mapping the spatial distribution of an attribute (sometimes when the attribute is smooth). It is used, however, for building conditional distributions for stochastic simulation

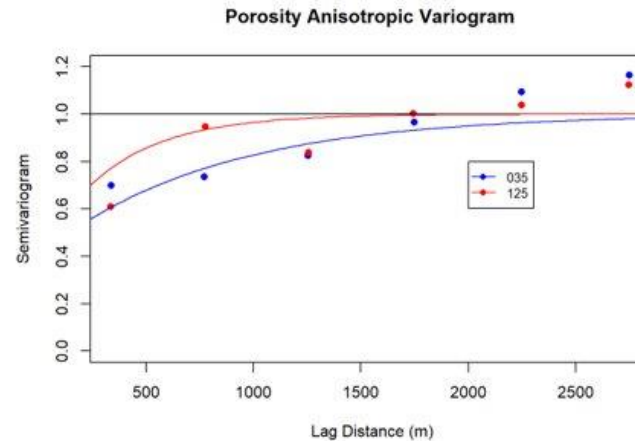
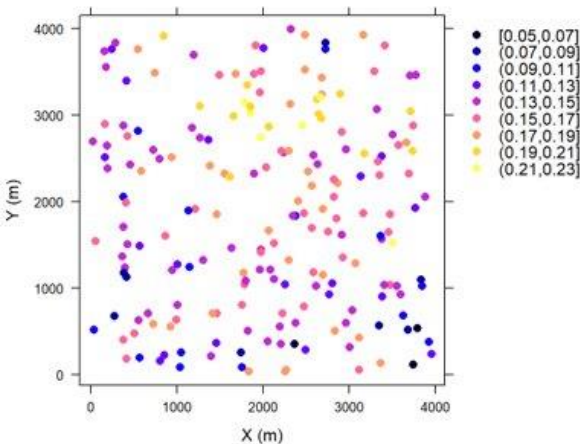
Kriging Demo in R

Markdown Tutorial on Geostatistical Spatial Estimation in R with gstat

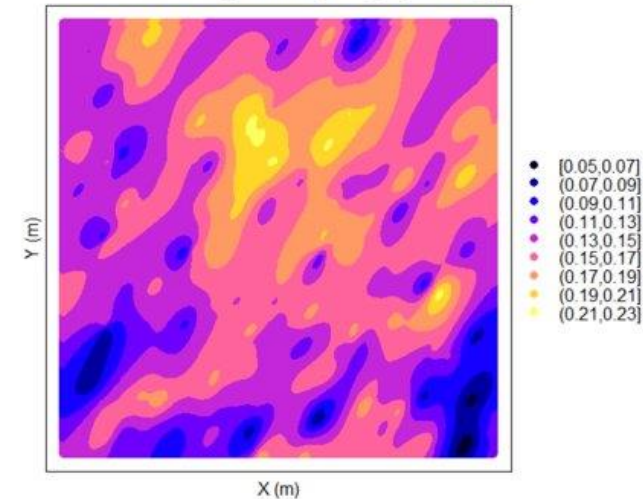
Michael Pyrcz, the University of Texas at Austin, @GeostatsGuy



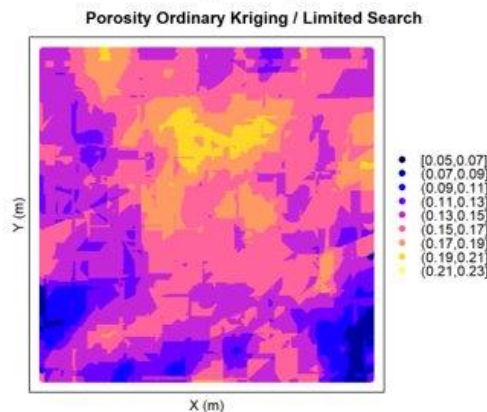
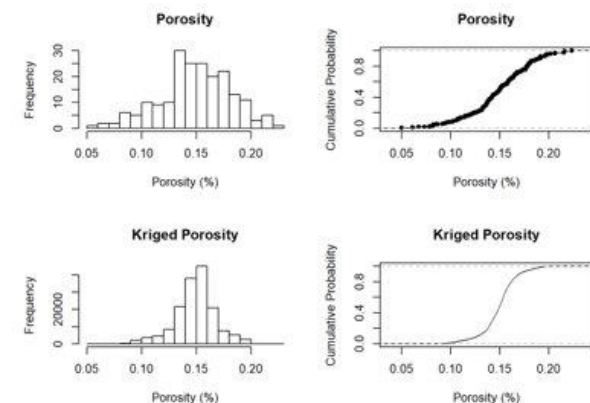
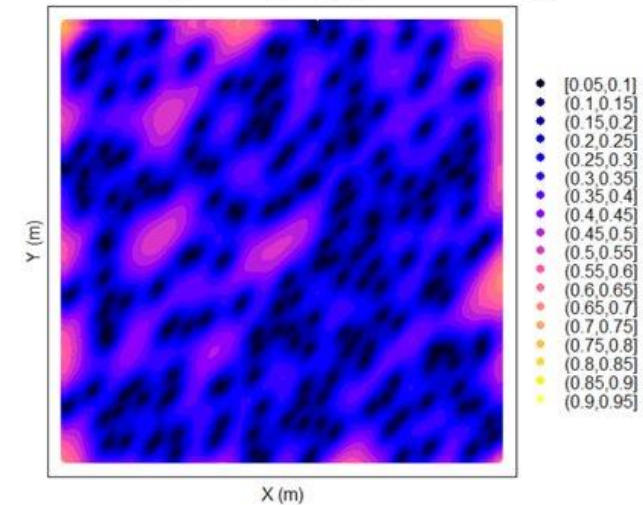
$$z^*(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha})$$



Porosity Ordinary Kriging No Nugget



Porosity Ordinary Kriging Variance No Nugget



Kriging Hands-on R

- At the very end of the demo, these lines modify the variogram model, rerun the ordinary kriging and plot the estimate and estimation variance.

```
# Let's remove the nugget effect and kriging again once again with unlimited search
# Anisotropic variogram with nugget effect removed
por.vm.ani.nonugget <- vgm(psill = 1.0, "Exp", 800, anis = c(0.35, 0.5), nugget=0.0)
por.vm.ani.nonugget

porosity.kriged.nonugget = krige(porosity~1, mydata, coords, model = por.vm.ani.nonugget, maxdist = Inf, nmin = 0, omax=Inf) # ordinary kriging
spplot(porosity.kriged.nonugget["var1.pred"], main = "Porosity Ordinary Kriging No Nugget", key.space = "right", cuts = cuts, xlab = "X (m)", ylab = "Y (m)")
spplot(porosity.kriged.nonugget["var1.var"], main = "Porosity Ordinary Kriging Variance No Nugget", key.space = "right", cuts = cuts.var, xlab = "X (m)", ylab = "Y (m)")
```

Some ideas to explore:

- Set the nugget high (use nugget parameter and set psill to 1-nugget).
- Set the range lower and higher (currently 800) in the function call.
- Increase and decrease the anisotropy ratio (currently 0.5) in the function call.

Review of Main Points

- Simple kriging (SK) is linear regression with some special properties:
 - Gives the mean and variance of conditional normal distribution
 - Best linear estimate for mean squared error criterion and variogram model
- Estimation variance is expected squared difference between estimate and truth that accounts for:
 - Initial variance if no data are available, the stationary variance of the property
 - The redundancy between the data
 - The closeness of the data to what is being estimated
- We derive simple kriging to minimize the error variance in expected value
- The use of SK estimates directly is somewhat limited, but it is used extensively under a multivariate Gaussian model for inference of conditional means and variances
 - We will discuss more next about simulation.

Spatial Estimation New Tools

Topic	Application to Subsurface Modeling
Trend + Residual Workflow	<p>Build a trend + residual workflow.</p> <p><i>Trend modeling communicating known vs. unknown variance.</i></p>
Trend Model	<p>Build a trend from available subsurface data.</p> <p><i>Use a moving window smoother or 1D vertical + 2D areal trend modeling method.</i></p>
Spatial Estimation	<p>Spatial estimation at an unsampled location.</p> <p><i>Use sparsely sampled subsurface data to calculate an estimate and estimation variance at an unknown location.</i></p>
Spatial Uncertainty	<p>Assess spatial uncertainty</p> <p><i>Use local kriging variance and maps of kriging variance to assess spatial uncertainty.</i></p>

Spatial Modeling with Geostatistics



Spatial Estimation

Lecture outline . . .

- Trend + Residual Method
- Trend Modeling
- Spatial Estimation

Prerequisites

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

Spatial Estimation

Stochastic Simulation

Uncertainty Management

Machine Learning